Temporal Build-Up of Psychoacoustic Roughness

Niclas A. Janßen[1], Arne Oetjen[2], Reinhard Weber[2], Steven van de Par[2]
[1] *Hearing Systems Group, Technical University of Denmark, DK-2800 Lyngby, Denmark, Email: najan@elektro.dtu.dk*
[2] *AG Akustik, Dept. Medizinische Physik und Akustik, Carl von Ossietzky Universität, 26111 Oldenburg, Deutschland*

Abstract

Psychoacoustic roughness occurs in sounds of rotating machinery and speech and has influence on different sound quality aspects like perceived pleasantness. Therefore, it has many applications in product sound design, e.g. for vehicle interior sounds. It arises from amplitude modulations within a certain frequency range of about 15 to 300 Hz. Despite variations of such modulations over time, roughness can occur in transient sounds, like when a car is accelerating. So far, roughness perception has rarely been studied with a focus on such temporal changes. One may question whether there exists a temporal build-up for the strength of perceived subjective roughness. In the present study, a roughness matching experiment for stimuli containing amplitude modulated parts with different lengths was conducted. It was found that the amount of perceived overall roughness seems to increase with increasing duration of the rough portion inside a sound with fixed overall length. From this, a general description for changes in roughness perception with such time-changes has been found and this is a basic approach for further modeling is discussed.

Experimental Method and Procedure

This study's aim is to investigate a build-up of psychoacoustic roughness with an increasing time-related amount of modulation in sounds. Thus, the crucial parameter is the duration of a modulated sound. Since roughness depends on the sound pressure level (SPL), it has to be kept the same for the modulated part and to prevent audible clicks in the stimuli, the phase of the modulated part is matched to the amplitude of the non-modulated part at the transition points between both. The modulated parts are placed in the middle of the stimuli sounds after a preliminary experiment showed no effect of placement towards either the beginning or end, compared to the middle (all ratings were the same within the error range). This small investigation was carried out in the same way as the following experiments.

To judge the perceived overall roughness, listeners are asked to compare the overall roughness of the stimulus A and another sound B of the same type and duration, which is modulated from beginning to end with the same frequency but an adjustable degree of modulation. Listeners then must adjust the sound B to the same overall roughness as the sound A, using a hardware slider.

In order get a general description of the roughness perception, different types of sounds are used in the experiments. The basic types of sounds are 500 Hz tones and broad-band pink noise (50 Hz to 22050 Hz). The choice of 500 Hz tones comes from vehicular acoustics and is not too far from the more often used 1000 Hz tones and still in the range of speech and music. Broad-band pink noise is closer to the spectrum of naturally occurring sounds ([1] and [2]). These basic sounds then have to be modulated with suitable frequencies, spanning relatively evenly over the range where rough sensations are evoked (as given in [2]). Therefore, modulation frequencies of 25; 50 and 75 Hz were selected for the experiment. The 25 Hz frequency can be related to modulations resulting from car engines ([1], [2]). Suitable modulation durations were found in listening tests. For the experiment durations from 80 ms to 4000 ms were selected.

The experiment started with participants receiving printed instructions, after which they are introduced to examples of rough and non-rough sounds and the setup they will be using. To ensure participants reassess the roughness in every stimulus, the slider input is mapped slightly different along its pathway with every stimulus. Additionally, changes on the slider should be perceived linearly. Thus an exponential scaling of the slider input is used, counteracting the logarithmic perception (cf. [2]). The program assures participants have listened to both signals before proceding. They are allowed to take a break at any time. To prevent introducing a trend to the results, the order of signals is permuted randomly for every participant (cf. [3]).

Measurements are carried out in a sound-isolated chamber (IAC type 40a), using Sennheiser HD 650 headphones driven by an RME Fireface USB sound card, which is set to a sampling frequency of 44100 Hz. Loudness is calibrated to be 65 dB sound pressure level (SPL). Between measurements regular checks are performed to assure the sound card delivers an correct output voltage amplitude and frequency.

Influence of Short Modulation Durations and the Number of Modulator Periods

The influence of duration (precisely the number of modulator periods) has been studied by Vogel in a paper in 1975 ([4]). There results show no evidence of an effect of duration on the perceived roughness. The number of listeners is not clearly listed for his experiment, but other experiments in the same paper list five or six participants. However, when the current authors listened to the same stimuli, the impression of roughness was not very strong. This lead to an investigation, following the type of experiment discussed before. 27 participants (12 female, 15 male), 14 experienced in psychoacoustic

experiments, had to rate a selection from the sounds Vogel used. Their age spanned from 14 to 43 years with a mean of 26.6 years. The stimuli were 2000 Hz tones with overall duration of 800 ms, degree of modulation of 0.8 and selected modulation frequencies of 50 Hz and 150 Hz, as used by Vogel. He based the durations of the modulated part on the number of modulator periods.
Almost all of 27 participants did rate the roughness lower in comparision to the stimuli used in [4]. It can only be theorized where the deviation comes from. Apart from the huge change in technology from 1975, a possible explanation for why Vogel's results deviate might be his participants may have rated the roughness of the modulated part of the signals either alone or very dominantly weigthed instead of the rating the overall roughness evenly over the complete stimulus. From the results it could be concluded that the duration and not the number of periods of the modulator influences the roughness perception, because the modulation frequency does not lead to a different shape of the roughness build-up. Hence, further experiments including higher durations were undertaken to show if this observation held for a wider range.

Experiment with Modulation Durations from 80 to 4000 ms

In this experiment, a roughness build-up with increasing duration of a modulated part shall be investigated. It is performed by 27 participants (13 female, 14 male), 15 experienced in psychoacoustic experiments. The variety of listeners should give a good picture of the average roughness perception and whether experience plays a role. The same applies to their age, which spans from 14 to 43 years with a mean of 26.7 years. The stimuli consist of 500 Hz tones and broad-band pink noise from 50 Hz to 22050 Hz with a overall duration of 4000 ms. In the middle, the stimuli contain a modulated part with durations varying from 80 ms to 4000 ms, modulated with a degree of modulation of 0.8 and frequencies of 25, 50 and 75 Hz. The total of 48 stimuli is measured twice and divided into eight sets of 12 stimuli. The total time for completing the measurement is about 100 minutes, excluding breaks.

Results

In this experiment the results are very similar for all the different signals, tone and modulation frequencies (cf. Fig. 1. An increase in roughness ratings from very low values up to the expected value for equal signals of 0.8 is visible. The general behaviour relative to the maximum roughness of the sound being modulated from beginning to end is shared among most participants and all the stimuli. Examples for individual ratings from the participants are given for tone and pink noise sounds in Fig. 2 for the modulation frequency of 75 Hz. Individual results for the other stimuli are very similar and therefore not depicted here.

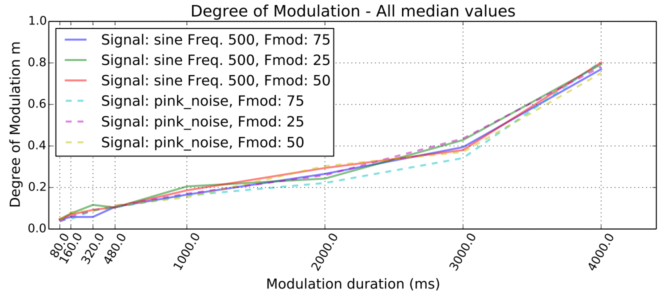

Abbildung 1: Median values for the degree of modulation (ordinate) from all measurements on the influence of the modulated part's duration on roughness perception (abscissa, 80 to 4000 ms). The stimuli are 500 Hz tones as well as broadband pink noise (50 to 22050 Hz), modulated for the given duration with frequencies of 25 Hz (solid green and dashed violet lines), 50 Hz (solid red and dashed yellow lines) and 75 Hz (solid blue and dashed cyan lines) and a degree of modulation of 0.8. Errorbars were omitted for clarity (about ±0.15).

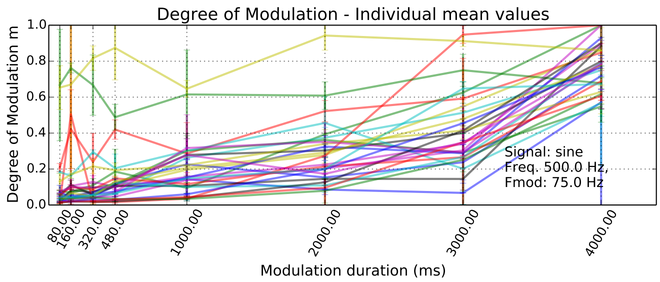

Abbildung 2: Individual participants' values for the degree of modulation (ordinate) from a measurement on the influence of the modulated part's duration on roughness perception (abscissa, 80 to 4000 ms). The stimuli consist of 500 Hz tones, modulated for the given duration with a frequency of 75 Hz and a degree of modulation of 0.8. The participants are marked by colored lines and the standard error is plotted with error bars on the measurement points.

Discussion of the Results

In these results a generally increasing roughness rating is shared among most participants and all the stimuli. Roughness perception appears to be much lower than expected when a modulated part is embedded in a sound. The theoretical alternative to a roughness saturation, longer durations for the modulated part were always perceived more rough, does not seem convincing.
Results show a range of rating behaviour (see Fig. 2 for examples of individual ratings for 75 Hz modulation frequency). A Shapiro-Wilks hypothesis test on the data for each stimulus condition shows the data is not always normally distributed. The Kurskal-Wallis one-way ANOVA is non-parametric ([5]) and hence does not rely on a normal distribution. It indicates a small difference in the adjusted degree of modulation for tone and noise signals (propability about 10.92 % for both coming from the same distribution). Further it shows two participants

with significantly higher average ratings. However, the representation of the data's median values is influenced to a lesser extent by the extreme ratings than the mean, thus reflecting the majority's behaviour better. Thus an exclusion of the two participants with such very high average ratings would not lead to distinctly different results and hence is omitted.

Analysing the Roughness Increase

How could the results be described and eventually modeled? For a modeling of the results, further experiments are required, but from current results one can identify a starting point. First, for further analysis the data is rescaled with the power law in order to have a roughness equivalent measure ($R \approx C * m^{1.6}$ with C as a constant dependend on the signal, [2]). Because of the minor difference between the types of sounds, the build-up of roughness could likely be described in the same way for these. For more complex sounds, duration, the experiments' independent variable, is not a suitable basis, if one aims for a general description or modeling. The output of auditory modulation filters could be a universal measure to assess the amount of modulation, as introduced in previous studies modeling auditory processing, especially with regard to modulation processing ([6] and [7]). This might be a way to model how listeners are processing sounds internally and how modulation information is extracted, upon which the roughness percept is based ([2]). The signals first have to pass a gammatone filterbank, modeling the frequency selectivity of the cochlea. After that the Hilbert-envelope of the output can be fed through additional modulation filterbanks per Gammatone filter channel. The relevant modulation information is contained in the envelope signals (see [8] and [6]). A Matlab implementation of the models is used with the standard settings but with the filter frequencies on the modulation frequencies used here[1]. The sound's modulation power P_{mod} at the filter output, appears to be a good choice for the model, similar to [7]. It is not dependent on the signal's duration but normalized to it and therefore more universally usable than the energy. The correction for the mean value of the signal is necessary to make it less dependent on statistical fluctuations (like with noise signals). The measure is calculated as follows:

$$P_{mod} = \frac{1}{D} * \sum_{i=1}^{N} |x_i - \mu|^2 \qquad (1)$$

Where μ is the mean value of the signal with N samples and duration D. The signal's duration can also be calculated from the length N and sampling frequency f_s: $D = N/f_s$.

Yet, even when there is no modulation added to the sounds, the calculated modulation power is not zero. The noise's inherent statistical fluctuation, leading to a modulation power, has to be taken into account. It has to be subtracted, for a corrected power measure: $P^C_{mod} = P_{mod} - P_0$. The power P_0 can be calculated using a sufficiently long, non-modulated sample of the sounds (see Tab. 1).

[1]It is available in form of the Auditory Modelling Toolbox, available at: http://amtoolbox.sourceforge.net/, Version 0.9.6 from 2014-06-25; functions 'gfb analyser new' and 'modfilterbankepsm'.

Tabelle 1: Modulation energy and power measures calculated from the 500 Hz tone and pink noise stimuli modulated with 75 Hz for the given duration (rounded), with the respective power P_{mod} and the corrected power P_c of the signal at the output of the modulation filter banks.

Dur. / ms	Tone		Noise	
	P_{mod} a.U.	P^C_{mod} a.U.	P_{mod} a.U.	P^C_{mod} a.U.
80	0.421	0.00841	0.0370	0.000739
160	0.424	0.0115	0.0409	0.00471
320	0.440	0.0273	0.0495	0.0133
480	0.432	0.0201	0.0581	0.0219
1000	0.477	0.0651	0.0876	0.0514
2000	0.541	0.128	0.150	0.114
3000	0.630	0.217	0.221	0.185
4000	0.685	0.273	0.296	0.260

Since the relative roughness increase with the modulation power in a sound always seems to follow the same curve, normalized power measures may be universally applicable, using the respective (maximum) corrected power measure for a sound completely modulated from beginning to end. To describe the results with a function, the strong increase in the roughness perception towards the highest duration can be seen as a saturation into the maximum roughness for that sound. The varying threshold for this maximum roughness perception can be modeled using a logistic function (cf. [9]). This can be combined with a linear function for the lower region:

$$R(P^C_{mod}) = \Delta_R * P^C_{mod} + \frac{R_{max} - (\Delta_R * P_{cross})}{1 + \mathrm{e}^{-\Delta_L * (P^C_{mod} - P_{shift})}} \qquad (2)$$

Where Δ_R and Δ_L determine the slope for the linear roughness increase, respectively the logistic transition, R_{max} is the maximum roughness. Which here must be equal to $0.8^{1.6}$, respecting the power law relation. It has the power $P^C_{mod,max}$, which can be found in the respective tables here (Tab. 1). Further, P_{cross} describes the crossover power level, where the model function transits into the logistic behaviour describing the threshold. P_{cross} is assumed to be around 3/4 of the maximum power $P^C_{mod,max}$ andP_{shift} to be about $7/8 * P^C_{mod,max}$. A more precise position should be determined in further experiments. Deviations of P_0 are allowed to be smaller than $\approx 10^{-3}$, described by P^C_0. The results can be found in Tab. 2 and Fig. 3.

The fit for the tones seems to follow the data points closely, confirmed by the respective values for the standard deviation of the fit parameters, the one for the pink noise sounds does not. Judging from these, the parameters determined for the tonal sounds fall within the error range of the ones determined for the noise sounds - and might therefore apply to these as well. An additional

Tabelle 2: Parameters for the given approximation of the roughness build-up according to eq. 2.

Parameter	Tone		Noise	
	Value	Std. dev. σ	Value	Std. dev. σ
Δ_R	9.85e-1	5.34e-2	8.62e-1	1.23e-1
P_0^C	-6.94e-3	9.10e-4	3.76e-4	2.62e-3
Δ_L	5.46e+1	7.49	1.00e+2	4.26e+1

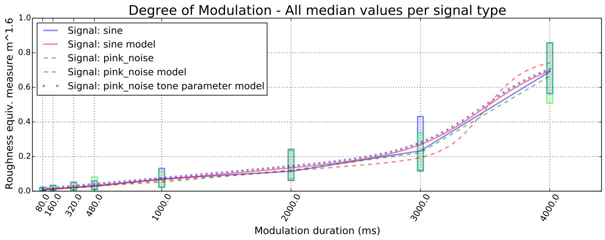

Abbildung 3: Fit for roughness build-up with the median values of the equivalent roughness measure (ordinate) from all measurements (abscissa, 80 to 4000 ms) per signal type. 500 Hz tone: Solid blue line, model solid red line; broad-band pink noise (50 to 22050 Hz): Dashed green line, model dashed pink line (data from modulation frequencies of 25, 50 and 75 Hz and a degree of modulation of 0.8). Shown are the 25 % and 75 % quantiles. Dottet violet line: Fit on the noise's corrected power measure with parameters determined from the tonal sounds.

graph plotted uses the parameters determined from the tonal sounds with the power values from the pink noise sounds. The result is a fit close to the actual data points. From this it can be concluded, the parameters determined from the tonal sounds seem to describe the overall behaviour regardless of the type of sound. These should thus be preferred. Also the parameters can be normalized to a maximum corrected modulation power value of 1, to make it more widely applicable. This is achieved by multiplying the parameters with the maximum power for the tone signals, P_{max}^C. They can be found in Tab. 3.

Tabelle 3: Normalized parameters for the given approximation of the roughness build-up according to eq. 2, parameters normalized to the maximum modulation power.

Parameter	Value	Std. dev. σ
P_{max}^C	1	-
P_0^C	0	-
Δ_R	2.56e-1	1.39e-2
Δ_L	14.2	1.95

Conclusions

Roughness is known to be influenced by level, modulation frequency and degree of modulation. Present results indicate the duration of a modulated sound has to be added to these measures; a build-up of overall roughness with duration of a modulated sound is noticeable. It begins rather low and increases linearly until a threshold is reached by which the maximum roughness sensation for a sound modulated from beginning to end is reached. Hence, many sounds in reality may appear less rough than one might expect. This is consistent for all the different stimuli used (tones and broad-band pink noise), regardless of modulation frequency and follows the same shape. Results can be described using a modulation power measure obtained from simulated auditory modulation filterbanks (cf. [6] and [7]), corrected for the modulation of the carrier signal, with a function consisting of linear and logistic parts. This could be a basis for a new model and prove useful for sound design. Models calculating the roughness of sounds need to be revised taking this into account. Calculating the roughness of signal blocks and using a percentile of that for the overall roughness will not represent the actual perception. It might be possible to hide or reduce the roughness of modulated components by limiting their duration. It needs to be determined if the findings of this study also apply to completely modulated sounds of varying duration and continuous sounds in further experiments, as there no context can be related to the modulated part.

Literatur

[1] P. Zeller. *Handbuch Fahrzeugakustik. Grundlagen, Auslegung, Berechnung, Versuch.* Vieweg und Teubner, Wiesbaden, 2013.

[2] E. Zwicker and H. Fastl. *Psychoacoustics: facts and models.* Springer, Berlin Heidelberg, 2007.

[3] J. HellbrÃ¼ck. *HÃ¶ren. Physiologie, Psychologie und Pathologie.* Hogefe, GÃ¶ttingen, 1993.

[4] A. Vogel. Ãœber den zusammenhang zwischen rauigkeit und modulationsgrad. *Acoustica*, 32:300–306, 1975.

[5] J. Bortz, G. A. Lienert, and K. Boehnke. *Verteilungsfreie Methoden in der Biostatistik.* Springer, Berlin Heidelberg New York, 1990.

[6] T. Herzke and V. Hohmann. Improved numerical methods for gammatone filterbank analysis and synthesis. *Acta Acoustica*, 2007.

[7] S. Joergensen and T. Dau. Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *Journal of the Acoustic Society of America*, 134:1475â“1487, 2013.

[8] D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet. *Auditory Signal Processing: Physiology, Psychoacoustics, and Models.* Springer, 2006.

[9] Manuel KÃ¼hner. *Haptische Unterscheidbarkeit mechanischer Parameter bei rotatorischen Bedienelementen.* Dissertation, Technische UniversitÃt MÃ¼nchen, MÃ¼nchen, 2014.